doi:10.1111/j.1365-2052.2004.01112.x

Characterization of a new BAC library for rainbow trout: evidence for multi-locus duplication

Y. Palti*, S. A. Gahr*, J. D. Hansen† and C. E. Rexroad III*

*USDA/ARS National Center for Cool and Cold Water Aquaculture, Kearneysville, WV, USA.†Center of Marine Biotechnology, University of MD Biotechnology Institutes, Baltimore, MD, USA.

Summary

A 10X rainbow trout bacterial artificial chromosome (BAC) library was constructed to aid in the physical and genetic mapping efforts of the rainbow trout genome. The library was derived from the Swanson clonal line (YY male) and consists of 184 704 clones with an average insert size of 137 500 bp (PFGE) or 118 700 bp (DNA fingerprinting). The clones were gridded onto 10 large nylon membranes to produce high-density arrays for screening the library by hybridization. The library was probed with 11 cDNAs from the NCCCWA EST project chosen because of interest in their homology to known gene sequences, seven known genes, and a Y-specific sex marker. Putative positive clones identified by hybridization were re-arrayed and gridded for secondary confirmation. FPC analysis of *Hind*III and *Eco*RV DNA fingerprinting was used to estimate the level of redundancy in the library, to construct BAC contigs and to detect duplicated loci in the semi-duplicated rainbow trout genome. A good correlation ($R^2 = 0.7$) was found between the number of hits per probe and the number of contigs that were assembled from the positive BACs. The average number of BACs per contig was 9.6, which is in good agreement with 10X genome coverage of the library. Two-thirds of the loci screened were predicted to be duplicated as the positive BACs for those genes were assembled into two or three different contigs, which suggests that most of the rainbow trout genome is duplicated.

Keywords BAC library, gene duplication, *Oncorhynchus mykiss*, physical mapping, rainbow trout.

Genome research for species of interest is facilitated by the development of species-specific tools such as well-characterized germplasm, physical and genetic mapping resources, large-insert libraries, public bioinformatic databases, and large quantities of sequence information. Rainbow trout (*Oncorhynchus mykiss*) is a model species for genome-related research activities focusing on aquaculture, carcinogenesis, toxicology, comparative immunology, disease ecology, physiology, transgenics, evolutionary genetics, and nutrition (Thorgaard *et al.* 2002). Current genomic resources available for rainbow trout research include multiple bacterial artificial chromosome (BAC) libraries (Katagiri *et al.* 2001, Phillips *et al.* 2003), clonal lines (Young *et al.* 1996), genetic maps (e.g. Nichols *et al.* 2003) and a large EST database (Rexroad *et al.* 2003; http://www.tigr.org/tdb/tgi/rtgi/).

Address for correspondence

Dr Yniv Palti, USDA/ARS National Center for Cool and Cold Water Aquaculture, 11876 Leetown Road, Kearneysville, WV 25430, USA.
E-mail: ypalti@ncccwa.ars.usda.gov

Accepted for publication 6 January 2004

Two rainbow trout BAC libraries representing 6.7X and 5.3X genome coverage with an average insert size of 58 kb and 110 kb, respectively, were previously reported (Katagiri *et al.* 2001). Another library representing 4.5X coverage and average insert sizes of 125 kb was constructed from genomic DNA of the OSU-142 XX female doubled haploid line (Young *et al.* 1996). As with all salmonids, rainbow trout experienced a recent genome duplication event resulting in a semi-tetraploid state (Allendorf & Thorgaard 1984). Therefore, the use of homozygous lines should be useful for detecting gene duplication in the process of constructing BAC contigs for physical mapping.

A 10X library was constructed by Amplicon Express (Pullman, WA, USA) from a Swanson YY doubled haploid male. The YY source for this library was selected to enable physical mapping and positional cloning of sex determination. The average insert size estimated from pulsed field gel electrophoresis (PFGE) of 100 *Not*I digested clones was 137.5 kb. The percent of clones that did not contain insert was 1.22% (654/53760), and 0.23% (416/184704) of the colonies had poor growth in the wells.

The library was screened with 19 radiolabelled probes. The template for eight probes was amplified by PCR from

Table 1 Number of positive hits, contigs assembled from the positive bacterial artificial chromosome (BAC)s, and BACs in each contig for the genes and ESTs used as probes in the BAC library screening.

Gene symbol	Gene (putative)	LG[1]	No. of hits	PCR verified	No. of contigs[2]	No. of BACs in each cntig	Reference	Forward primer	Reverse primer
MHC-I (HLA-A)	Major histocompatibility 1	3 and 16	40	40	3	14, 13, 9	Hansen *et al.* (1999)	CAGTGTCTCTGCTCCAGAAGG	TCAGAACCTCGATGAAGTCCTT
TAP2B	Transport and activating protein 2B	3 and 16	19	18	2	11 and 7	Hansen *et al.* 1999	ATGGCTGTAGGTCTATGCAT	TAGAAGATGAGGACTCTCATG
SCAR163	Chromosome Y marker	1	162*	13	1	10	Felip Edo *et al.* 2003	CTTCTG TCTACCAAAATC	CATCAAGTCACATGACTAAC
NPY	Neuropeptide Y		7	7	2	4 and 2	J. Silverstein (per. Comm.)	GTCCAGATATGATGAACCCTCGTT	CGTCGCCACGACGATGA
DAA (HLA-DRA)	Major histocompatibility II alpha chain	29	0	0	–	–	Genbank: AJ251432	GGAGTTGCTACCGGAACATT	CCTTTCTCAAACGGGTCTCTATCT
TAP1	ATP dependent transporter 1	27	16	15	1	13	Hansen *et al.* 1999	TTTGATAAACCAGACTCCTGTCGC	CCGTGCATGTGACTTGGACCAT
DAB (HLA-DRB)	Major histocompatibility II beta chain	29	44*	1	–	–	Palti *et al.* 2001	TACAGCGCCATACTGGACAA	TGAGCTCAGTCTGACATGGG
GH2	Growth hormone 2	2 and 9	19	18	2	13 and 4	GenBank: J03797	CAAGTGTCCTCTTCACGCA	GGTACTCCCAGGATTCAATCA
MRPS16	40S ribosomal protein S16		13	**	2	7 and 3	GenBank: CA369166	**	**
ZFP238	Homo sapiens zinc finger protein 238		22	**	2	18 and 2[3]	GenBank: CA372518	**	**
GNAI2	GTP-binding regulatory protein Gi		31	**	4 or 3[3]	6, 8, 4 and 5 or 6, 12 and 7[4]	GenBank: CA369058	**	**
HAMP	Hepcidin (Precursor)		35	**	2	26 and 3	GenBank: CA369704	**	**
ACO1	Iron regulatory protein 1		27	**	2	12 and 11	GenBank: CA369541	**	**
FABP	fatty acid-binding protein b		19	**	1	12	GenBank: CA369721	**	**
AMBP	alpha-1-microglobulin		7	**	2 or 1[4]	4 and 2, or 6[5]	GenBank: CA372733	**	**
CXC-R4	CXC chemokine receptor		11	**	1	11	GenBank: CA372676	**	**
ID1[6]	Inhibitor of DNA binding 1		26	19	–	–	GenBank: CA369758	CTTTGGAAACTACAGCTACACC	GTAGTCAATGACGTGCTGGA
ID6	Inhibitor of DNA binding 6		25	17	2	18 and 5	GenBank: CA372740	CTTTGGAAACTACAGCTACACC	GTAGTCAATGACGTGCTGGA
MYD118[6]	Myeloid differentiation 118		12	12	–	–	GenBank: CA369202	GGATTACTTCATTGACACGG	TTTGCAGACTCGTAGACTCC
Average:			28.2*	14.5	1.8	9.6[7]			

[1]Linkage group assignments are according to Nichols *et al.* 2003 (OSU x Arlee cross).

[2]FPC cutoff of e^{-14} (see methods for a detailed description).

[3]The two BACs in this contig have identical *Hind*III DNA fingerprints.

[4]FPC cutoff of e^{-13}.

[5]FPC cutoff of e^{-12}.

[6]BACs positive to these probes were not analyzed by FPC analysis for contigs assembly.

[7]The average was calculated for the minimum number of contigs per probe (i.e. assuming three contigs for *GTPBP-Gi* and one contig for AMBP)

*Average number of hits without *SCAR163* and *DAB* is 19.4.

**No PCR primers for these probes.

genomic DNA. Primer sequences are listed in Table 1. The specific genes were selected because they were previously mapped and characterized in rainbow trout. The other 11 probes were ESTs from the cDNA library we previously described (Rexroad *et al.* 2003). They were selected because of interest in their homology to known gene sequences. The probes were hybridized to high-density filters following established procedures (Sambrook *et al.* 1989). Each filter was hybridized with a cocktail of all the probes at 10^6 cpm per probe per ml. Positive clones were picked from the library and re-arrayed into 96 well plates. Secondary hybridization filters were gridded from the re-arrayed positive clones and screened with individual probes. Clones were re-arrayed again for PCR verification and for *Hind*III and *Eco*RV DNA fingerprinting. BAC DNA for the PCR verification and DNA fingerprinting was purified using the REAL 96 plasmid kit on a Biorobot 8000 (Qiagen, Valencia, CA, USA).

The average number of positive hits in the secondary screen was 28.2 (± 34.4) per probe and almost all of the positive BACs that were screened by PCR were confirmed to contain the sequence from which the probe was derived, with the exception of two probes (*DAB* and *SCAR*163). Without these two the average number of positive hits was 19.4 (± 10.6) per probe. The high number of false positives in those two probes is likely the result of cross hybridization to other loci in the genome. *DAB* contains an *IgSf* domain with sequence similarity to other cell surface receptors (e.g. *NITR*2; Yoder *et al.* 2002) and *SCAR*163 appears to contain a repetitive element.

A standard DNA fingerprinting protocol was followed (Marra *et al.* 1997). Gel images were analysed by Image 3.10b and FPC (fingerprinting contigs) V6.4 (Sulston *et al.* 1989; Soderlund *et al.* 1997). A fixed tolerance of seven was used with an initial cut-off value of e^{-14} and a bury value of 10%. The cut-off value was subsequently raised to e^{-13} and e^{-12} for expanding contigs that corresponded to specific genes. DNA fingerprinting bands were scored in the size range of 1.6–40 kb. The average number of bands per lane was 18.5 (± 5.4) in the *Hind*III gels and 17.2 (± 3.0) in the *Eco*RV gels. An average insert size of 118.7 kb with a standard deviation of 22.15 kb was estimated from the *Hind*III fingerprints of the 371 clones. Two hundred and fourteen of the 371 clones (58%) were found to be identical by *Hind*III FPC analysis. The range was 2–7 clones in a group of BACs with identical DNA fingerprints. A subsample of 45 *Hind*III identical clones that were positive to 5 different probes was analyzed by *EcoRV* FPC. Only two of the 45 (4.4%) were identical by their *Eco*RV fingerprints. This suggests that 2.6% (4.4% × 58%) of the BACs in the library are completely identical.

Twenty-seven contigs were constructed by FPC from BACs that were positive to 15 probes (Table 1). A good correlation ($R^2 = 0.7$) was found between the number of positive hits per probe and the number of contigs that were assembled from those BACs. The average number of BACs per contig was 9.6 (± 5.6). A list of the clones that were assembled to each contig (cutoff = e^{-14}), the insert size of each clone and the size of each contig can be downloaded from http://ncccwa.ars.usda.gov/New%20Public%20Information.htm (under Yniv Palti choose Appendix 1; open as a Microsoft Excel file).

The FPC analysis was useful for identifying duplicated loci in the trout genome. Rainbow trout *growth hormone* 2 (*GH*2) was used to test the utility of this resource for identifying duplicated loci as it was previously mapped to two homeologous linkage groups (Nichols *et al.* 2003). The 18 *GH*2 BACs identified in this study were assembled into two contigs, as expected from the genetic mapping data. Additionally, our FPC analysis indicated that the locus containing *MHC-I* and *TAP*2 is also duplicated. The duplication of this locus was recently confirmed by both linkage analysis and *in-situ* hybridization (Phillips *et al.* 2003). It is important to note that the MHC genes that were used as probes in this study (*MHC-I*, *DAB* and *TAP*1) were previously mapped to different linkage groups (Phillips *et al.* 2003), and therefore represent different loci of the rainbow trout genome. Another confirmation for the utility of our DNA fingerprinting analysis for detecting gene duplication was recently reported by Gahr *et al.* (2003), who identified two different transcripts for the *ID*6 gene in rainbow trout (AY325275, AY325276). Each of the two contigs assembled for that gene harbors one of the two transcripts, which indicates that *ID*6 is duplicated.

Our contig assembly analysis suggests that two-thirds of the trout genome loci are still duplicated. Eight of the 12 random loci we screened (excluding *GH*2, the Y-specific sex marker and treating *MHC-I* and *TAP*2 as one locus, respectively) were assembled into two contigs. This estimate is in good agreement with Nichols *et al.* (2003) that found homeologies among two-thirds of the linkage groups in the OSUxARL genetic map.

The assembly of positive BACs into three contigs was observed in two different instances (for the probes *MHC-I* and *GNAI*, respectively). The *MHC-I* results can be explained by locus duplication and cross-hybridization with other genes that possess sequence motifs similar to the specific probe used (*UBA* exon 4). Two of the *MHC-I* contigs were composed of BACs that were positive for *TAP*2*B* indicating that they are part of the two previously identified MHC class I regions (Phillips *et al.* 2003). The third contig is likely to contain genes that share similar sequence motifs with *UBA* exon 4, such as *CD*1, but may also represent local duplication of the *MHC-I* gene.

DAB and *DAA* are linked in rainbow trout as members of the MHC class II region (Phillips *et al.* 2003). Only one BAC clone was found to be positive for this locus in this library. No positive clones were found in the OSU library, which was screened four times with different probes for this locus. The two libraries were constructed with genomic DNA partially

digested with *Hind*III. It is possible that the MHC class II region of rainbow trout is over or under represented by *Hind*III sites.

BACs positive by hybridization and PCR to *SCAR*163, the Y-specific sex marker, were assembled into one contig indicating that the male sex determining locus is present in this library. This marker was found to be male-specific in the Swanson clonal line and the BACs identified here were shown by fluorescent *in situ* hybridization (FISH) to hybridize to the sex chromosome of rainbow trout (Felip Edo *et al.* 2003). This contig may be expanded in the future to initiate positional cloning of the sex determination gene in rainbow trout.

In this paper we described the construction, characterization and utilization of the new Swanson BAC library for rainbow trout. The library provides 10X genome coverage and an excellent tool for positional cloning of loci of interest and for integration of the genetic and physical maps. It can be used in conjunction with the other rainbow trout libraries mentioned here to generate a comprehensive physical map, and to construct a minimum tiling paths for region-specific or whole-genome sequencing. In addition, high density filters for probe hybridization screening and DNA super pools for PCR screening of this library are available and clones are distributed by a not for profit organization at cost.

This is the first report in which physical mapping tools are utilized for evaluating the level of genome duplication in salmonids.

Acknowledgements

The authors acknowledge Renee Fincham and Roseanna Athey for their technical assistance and thank Alicia Felip Edo and Gary Thorgaard for the *SCAR*163 sequence information. We also thank Gary Thorgaard for contributing the source of the genomic DNA for the library construction and Thomas Kocher for his review of earlier version of the manuscript. The research was supported in part by USDA-NRICPG grant number 2002-03472. Mention of trade names or commercial products in this publication is solely for the purpose of providing specific information and does not imply recommendation or endorsement by the U.S Department of Agriculture.

References

Allendorf F.W. & Thorgaard G.H. (1984) Tetraploidy and the evolution of salmonid fishes. In: *Evolutionary Genetics of Fishes* (ed. by B.J. Turner), pp. 1–46. Plenum Press, New York.

Felip Edo A., Fujiwara A. & Thorgaard G.H. (2003) Segregation of Y chromosome-specific molecular markers in rainbow trout by AFLP analysis: characterization of SCAR DNA fragment and cytogenetic studies in different male clonal lines. *Plant & Animal Genome XI*, San Diego, CA, USA, p. 9.

Gahr S., Keele J., Palti Y. & Rexroad C. (2003) Sequence analysis and utililization of rainbow trout expressed sequence tags (ESTs). *Plant & Animal Genome XI*, San Diego, CA, USA, p. 241.

Hansen J.D., Strassburger P., Thorgaard G.H., Young W.P. & Du Pasquier L. (1999) Expression, linkage, and polymorphism of MHC-related genes in rainbow trout, Oncorhynchus mykiss. *Journal of Immunology* **163,** 774–86.

Katagiri T., Asakawa S., Minagawa S., Shimizu N., Hirono I. & Aoki T. (2001) Construction and characterization of BAC libraries for three fish species; rainbow trout, carp and tilapia. *Animal Genetics* **32,** 200–04.

Marra M.A., Kucaba T.A., Dietrich N.L., Green E.D., Brownstein B., Wilson R.K., McDonald K.M., Hillier L.W., McPherson J.D. & Waterston R.H. (1997) High throughput fingerprint analysis of large-insert clones. *Genome Research* **7,** 1072–84.

Nichols K.M., Young W.P., Danzmann R.G., Robison B.D., Rexroad C., Noakes M., Phillips R.B., Bentzen P., Spies I., Knudsen K., Allendorf F.W., Cunningham B.M., Brunelli J., Zhang H., Ristow S., Drew R., Brown K.H., Wheeler P.A. & Thorgaard G.H. (2003) A consolidated linkage map for rainbow trout (Oncorhynchus mykiss). *Animal Genetics* **34,** 102–15.

Palti Y., Nichols K.M., Paulson K.I., Parsons J.E. & Thorgaard G.H. (2001) Association between DNA polimorphisms tightly linked to MHC class II genes and IHN virus resistance in backcrosses of rainbow and cutthroat trout. *Aquaculture* **194,** 283–89.

Phillips R., Zimmerman A., Noakes M., Palti Y., Morash M., Eiben L., Ristow S.S., Thorgaard G.H. & Hansen J.D. (2003) Physical and genetic mapping of the rainbow trout major histocompatibility regions: evidence for duplication of the class I region. *Immunogentics* **55,** 561–69.

Rexroad C.E. III, Lee Y., Keele J.W., Karamycheva S., Brown G., Koop B., Gahr S.A., Palti Y. & Quackenbush J. (2003) Sequence Analysis of a Rainbow Trout cDNA Library and Creation of a Gene Index. *Cytogenetics and Genome Research*, in press.

Sambrook J., Fritsch E.F. & Maniatis T. (1989) *Molecular Cloning: A Laboratory Manual.* 2nd edn. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.

Sulston J., Mallett F., Durbin R. & Horsnell T. (1989) Image analysis of restriction enzyme fingerprint autoradigrams. *Computation and Application Bioscience* **5,** 101–06.

Soderlund C., Longden I. & Mott R. (1997) FPC: a system for building contigs from restriction fingerprinted clones. *Computation and Application Bioscience* **13,** 523–35.

Thorgaard G.H., Bailey G.S., Williams D., Buhler D.R., Kaattari S.L., Ristow S.S., Hansen J.D., Winton J.R., Bartholomew J.L., Nagler J.J., Walsh P.J., Vijayan M.M., Devlin R.H., Hardy R.W., Overturf K.E., Young W.P., Robison B.D., Rexroad C.E. & Palti Y. (2002) Status and opportunities for genomics research with rainbow trout. *Comparative Biochemistry and Physiology Part B* **133,** 609–46.

Yoder J.A., Mueller M.G., Nichols K.M., Ristow S.S., Thorgaard G.H., Ota T. & Litman G.W. (2002) Cloning novel immune-type inhibitory receptors from the rainbow trout, Oncorhynchus mykiss. *Immunogenetics* **54,** 662–70.

Young W.P., Wheeler P.A., Fields R.D. & Thorgaard G.H. (1996) DNA fingerprinting confirms isogenicity of androgenetically derived rainbow trout lines. *Journal of Heredity* **87,** 77–81.